

The Choir Invisible: zur Analyse der gesundheitsbezogenen Panelmortalität im Sozio-Ökonomischen Panel SOEP

Schnell, Rainer; Heller, Günther

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

SSG Sozialwissenschaften, USB Köln

Empfohlene Zitierung / Suggested Citation:

Schnell, R., & Heller, G. (2000). The Choir Invisible: zur Analyse der gesundheitsbezogenen Panelmortalität im Sozio-Ökonomischen Panel SOEP. In U. Helmert, K. Bammann, W. Voges, & R. Müller (Hrsg.), *Müssen Arme früher sterben? Soziale Ungleichheit und Gesundheit in Deutschland*, (S. 115-134). Weinheim: Juventa Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-121750>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

The Choir Invisible

Zur Analyse der gesundheitsbezogenen Panelmortalität
im Sozio-Ökonomischen Panel (SOEP)

Einführung

Neben der Inkonstanz von Messinstrumenten und Paneffekten gilt die Panelmortalität als das zentrale Problem bei der Durchführung und Beurteilung von Panelstudien (vgl. Kasprzyk et al. 1989, Pol 1989, Rendtel 1995a). Panelmortalität verringert zunächst den Umfang der untersuchten Stichprobe. In der Regel ist diese Verringerung beträchtlich; z.B. befanden sich beim Sozio-Ökonomischen-Panel (SOEP) nach 12 Jahren und 11 Wellen weniger als 50% der Ausgangsstichprobe noch in der Studie (SOEP-Info 1998). Das größere Problem der Panelmortalität besteht darin, dass eventuelle systematische Mechanismen, die zu Ausfällen führen, sowohl Punktschätzer interessierender Parameter als auch die Ergebnisse kausaler Analysen verzerren können. Eine solche Verzerrung wird um so gravierender erscheinen, wenn die mit dem Ausfallprozess verbundenen Variablen zentrale Themen der aktuellen Analyse sind, z. B. Gesundheitsvariablen in einem Gesundheitspanel.

Dabei kann von zwei unterschiedlich gerichteten Prozessen bei Vorliegen von Krankheit ausgegangen werden: Einerseits ist zu erwarten, dass erkrankte Personen eher verweigern, bzw. eher stellvertretend für sie verweigert wird („Unsere Oma ist zu krank, um an Ihrer Untersuchung teilzunehmen“), andererseits kann angenommen werden, dass zumindest bei Vorliegen von nicht allzu schwerwiegenden bzw. lebensbedrohlichen Erkrankungen die Erreichbarkeit der Zielperson höher ist. Insgesamt scheinen aber meist erhöhte Nonresponse-raten durch Krankheit zu resultieren (Moum 1994, Cheesbrough 1993)

Geht man z. B. von der Annahme aus, dass erhöhte Morbidität zu einer Erhöhung der Ausfallwahrscheinlichkeit führt und Morbidität in der Population derart schichtspezifisch verteilt ist, dass Personen unterer sozialer Schichten eher erkranken, dann wäre zu erwarten, dass der Gesundheitszustand der Bevölkerung durch Befragungen überschätzt wird. Weiterhin würde ein derartiger Ausfallmechanismus kausale Längsschnittanalysen gefährden. Z. B. könnte eine vermeintlich festgestellte abnehmende Prävalenz chronisch psychischer Erkrankungen durch eine überproportionale Panelmortalität der chronisch psychisch Kranken aus der Grundschrift bedingt sein.

In Deutschland existiert kein Gesundheitspanel, das auf einer echten Zufallsstichprobe der allgemeinen Bevölkerung basiert¹. Allerdings enthält das SOEP als bedeutendste Panelerhebung der allgemeinen Bevölkerung in der BRD (Hanefeld 1987) auch eine ganze Reihe von Gesundheitsfragen. Darüber hinaus wird das SOEP in zunehmenden Maße auch als Datenbasis für Studien aus dem Gesundheits- und „Public Health“ Bereich genutzt². Beide genannten Punkte ließen es sinnvoll erscheinen, eine Analyse der gesundheitsbedingten Panelmortalität mit dem SOEP durchzuführen³. Über erste Ergebnisse vorbereitender Analysen wird im Folgenden berichtet.

Datenbasis

Das SOEP ist eine Longitudinalstudie, die auf mehrstufigen Zufallsstichproben aus den in Deutschland lebenden Personen basiert (Westdeutschenstichprobe, Ausländerstichprobe, später auch Ostdeutschenstichprobe, Zuwandererstichprobe, Auffrischungstichprobe)⁴. Die Nettostichprobe enthielt in der ersten Welle 1984 insgesamt 16.205 Personen. Davon konnten 12.245 interviewt werden. Ein 13-Jahreslängsschnitt war für knapp mehr als 40% der Personen (5.030) möglich (Pannenberg 1998, Pannenberg et al. 1998, Pannenberg 1997, Pannenberg & Rendtel 1996, Rendtel 1995b, Soep-Info 1998).

Im SOEP wurden insgesamt 38 Variablen zur Gesundheit erhoben. Nach inhaltlichen Kriterien können dabei Maße der subjektiven Gesundheitseinschätzung (z.B. Zufriedenheit mit der Gesundheit) und Maße der objektiven gesundheitlichen Beeinträchtigung (z. B. Schwerbehinderung oder Pflegebedürftigkeit) bzw. der Inanspruchnahme medizinischer Dienstleistungen (z.B. Anzahl der Arztbesuche) unterschieden werden. Wie eine Übersicht über die Erhebungszeitpunkte der Indikatoren zeigt, kommen aber für eine Analyse nur wenige Variablen in Frage, da alle anderen Indikatoren nur in wenigen Wellen erhoben wurden. Von den verbleibenden Indikatoren wurden die Variablen *Arztbesuch*, *Krankenhausaufenthalt*, *Gesundheitszufriedenheit*, *Schwerbehinderung* und *Pflegebedürftigkeit*⁵ für die Analysen ausgewählt.

¹ Die einzige Ausnahme stellt eine einmalige Wiederholungsbefragung eines Querschnitts des Nationalen Untersuchungssurveys, initiiert durch das Deutsche Institut für Bevölkerungsforschung, dar.

Zum Konzept der „allgemeinen Bevölkerung“ und dem Umfang der (auch für medizinische Untersuchungen) relevanten Subpopulationen, die durch dieses Konzept ausgeschlossen werden, vgl. Schnell 1991.

² Die Datenbank „SOEPLIT“ (zugänglich über „WWW.DIW.DE“) enthielt im November 1998 bereits 36 Publikationen zum Stichwort Gesundheit.

³ Die in dieser Veröffentlichung verwendeten Daten des Sozio-ökonomischen Panels (SOEP; Version GSOEP12) wurden vom Deutschen Institut für Wirtschaftsforschung (DIW), Berlin, bereitgestellt.

⁴ Seit 1990 auch der in den neuen Bundesländern lebenden Personen.

⁵ Im Datensatz PFLEGE findet sich die Variable *maxvar* in der die Pflegebedürftigkeit in verschiedenen Schweregraden festgelegt ist. Im gleichen Datensatz findet sich die

Darüber hinaus wurde ein *Krankheitersatzindex* (KEI) (Fuchs und Hansmeier 1996) verwendet. Der KEI soll den allgemeinen Gesundheits- bzw. Krankheitszustand wiedergeben. Der Index wurde von Fuchs und Hansmeier aus gesundheitsbezogenen Angaben im SOEP entwickelt und mit Daten des Nationalen Untersuchungs Surveys (NUS) validiert. Innerhalb des NUS (Erhebungszeitpunkt: 1985) korreliert der KEI gut mit relevanten Krankheits- bzw. Beschwerdeangaben.

Der KEI für die Person i ist wie folgt definiert:

$$KEI_i = \text{Anzahl Arztbesuche}_i * \text{Gewicht} + \text{Anzahl Krankenhausnächte}_i$$

Der Gewichtungsfaktor wird für nicht chronisch Kranke auf 1 gesetzt. Für chronisch Kranke wird folgender Faktor berechnet⁶:

$$\text{Gewicht chronisch Kranke} = (A_c/N_c)/(A_g/N_g)$$

wobei

A_c	=	Summe aller Arztbesuche chronisch Kranker in der untersuchten Welle
N_c	=	Anzahl chronisch Kranker in der untersuchten Welle
A_g	=	Summe aller Arztbesuche aller Gesunden in der untersuchten Welle
N_g	=	Anzahl der Gesunden in der untersuchten Welle

Von entscheidender Bedeutung für jede Analyse von Ausfällen ist die Definition der Ausfallursachen. Bei Ausfällen in Bevölkerungsumfragen (Nonresponse) wird traditionell lediglich zwischen „Schwer-Ereichbaren“ (Not at homes), „Nicht-Befragbaren“ (Unable to answers) und „Verweigerern“ (Refusals) unterschieden (Cochran 1972, Schnell 1997). Die Daten des SOEP geben weit detailliertere Informationen über das Feldgeschehen. In der Datei PPFAD liegen die Informationen zur Teilnahme bzw. Nichtteilnahme (Variable *\$netto*) vor⁷. Diese Variable gibt den Befragungsstatus jedes Panelteilnehmers für jede Welle wieder und erlaubt die Unterscheidung, ob er

- in die Stichprobe gehört und einen Personenausfall darstellt (= 0),
- befragt wurde (= 1),
- ein Kind unter 16 Jahren ist und somit nicht befragt wurde (= 2),
- nur im Adreßbuch verzeichnet war, aber kein Kontakt zu ihm oder zum Haushalt zustande kam (kein Kontakt mit Haushalt = 3),

Angabe in welchen Jahr die Pflegebedürftigkeit erhoben wurde. Aus diesen Angaben wurde eine Variable für Pflegebedürftigkeit berechnet.

⁶ Darüber hinaus nehmen Fuchs und Hansmeier noch eine u. E. nicht ganz unproblematische Mittelwertersetzung für diejenigen chronisch Kranken vor, die keinen Arztbesuch in den letzten 3 Monaten angaben. Hier wurde daher ein in dieser Hinsicht leicht modifizierter Index berechnet.

⁷ Das Symbol „\$“ wird benutzt um eine der möglichen Wellen (A, B, C usw.) zu bezeichnen.

- zu einem späteren Zeitpunkt als Lücke nacherhoben wurde (= 4),
- wegen Umzug ins Ausland oder Tod nicht mehr zur Stichprobe gehört (trifft nicht zu bzw. kein Personeneintrag = -2; Personeneintrag in der Datei PPFAD gelöscht = -3)

Die Variable *\$perg* in der Datei \$PBRUTTO enthält darüber hinaus die Bearbeitungsergebnisse für alle Personen. Dabei werden die folgenden Ausprägungen unterschieden⁸:

- teilweise realisiert (= 0)
- realisiert (= 1)
- derzeit nicht durchführbar (= 2)
- derzeit nicht teilnahmebereit (= 3)
- endgültig verweigert (= 4)
- ins Ausland verzogen (= 5)
- verstorben (= 6)
- verzogen, nicht weiterverfolgt (= 7)
- in Feldzeit nicht gefunden (= 8)
- Adresse, Haushalt nicht auffindbar (= 9)

Da das SOEP als Haushaltspanel konzipiert ist liegen entsprechende Angaben auch für die Haushalte vor (*\$hergs* bzw. *\$hergsz* in der Datei \$HBRUTTO)⁹. Da Krankheit und Gesundheit zunächst ein individuelles Merkmal und keine Eigenschaft von Haushalten darstellt, wurde für die folgende Analyse das Bearbeitungsergebnis der Person (*\$perg*) als Kriterium für Teilnahme bzw. Non-response verwendet¹⁰.

⁸ Darüber hinaus liegen auch zweistellige Kodierungen der Ausfallursachen vor (*\$pergz*), die genauere individuelle Ausfallmusteranalysen erlauben würden. Solche Analysen wurden bislang unseres Wissens nicht publiziert. Auf Grund des immensen Aufwands wurde auch hier auf solche Analysen verzichtet.

⁹ Zusätzlich liegen für die Wellen B-H Angaben über das Befragungsergebnis aus dem Vorjahr vor (*\$hergv*, *\$pergv*), die in wenigen Fällen nicht mit den Angaben aus den Vorjahren übereinstimmen. Für die folgenden Analysen wurde immer auf die Information des aktuellen Bearbeitungsjahres zurückgegriffen.

¹⁰ Allerdings muss ergänzt werden, dass die Variable *\$perg* nicht mehr erhoben wird, wenn durch die Vorarbeiten im Feld klar wird, dass diese Person nicht mehr zur Studienpopulation gehören (Wegzug ins Ausland oder Tod). In diesem Fall weist diese Person im File YPBRUTTO in der Variable *ypzug* den Grund für den Ausfall und in der Variable *erhebj* das Jahr des Ausfalls aus. Die Variable *\$netto* (in PPFAD) ist für diesen Teilnehmer den Wert -2 gesetzt. Auch diese Personen wurden zunächst in den hier berichteten Analysen berücksichtigt. Bei dem Versuch der Replikation der Arbeiten der DIW Arbeitsgruppe fand diese Personengruppe aber keine Berücksichtigung.

Bisherige Analysen des Nonresponse im SOEP

Den Autoren sind drei verschiedene Analysen des Nonresponse im SOEP bekannt¹¹. Keine der Analysen untersucht explizit den möglichen Effekt von Krankheit oder Gesundheit auf die Panelmortalität¹².

Esser et al. (1989) berichten in einer Analyse der ersten 4 Wellen des SOEP systematische Panelmortalität an Hand der Variablen Bildung, Erwerbstätigkeit und Stellung im Beruf. Allerdings wird lediglich die Veränderung des Anteils einer bestimmten Berufsgruppe im Laufe der Zeit berichtet. Eine Aufgliederung nach unterschiedlichen Ausfallursachen oder eine Kontrolle von Drittvariablen bei der Erklärung der Ausfälle erfolgt nicht.

In einer neueren Analyse führen Engel und Poetschke (1997) anhand des Merkmals „Arbeitslosigkeit“ erweiterte loglineare Modelle über vier Wellen durch und berichten, dass ab der dritten Welle die Ausfälle nicht mehr systematisch, sondern nach zufälligem Muster erfolgen. Dabei muss allerdings einschränkend erwähnt werden, dass neben Arbeitslosigkeit keine weiteren möglichen Einflussfaktoren auf die Panelmortalität betrachtet werden und zur Interpretation nur globale Fitmaße für das Gesamtmodell herangezogen werden. Schließlich wird bei diesen Modellen auch nicht nach unterschiedlichen Ausfallursachen unterschieden.

Alle anderen Analysen zum Nonresponse im SOEP stammen aus der SOEP Arbeitsgruppe des DIW und basieren auf den Arbeiten von Ulrich Rendtel (vgl. z.B. Pannenberg 1998, Pannenberg 1997, Pannenberg et al. 1998, Pannenberg & Rendtel 1996, Pischner 1994, Rendtel 1995a, Rendtel 1995b). Bei diesen Arbeiten wird zwischen Ausfällen durch Kontaktverlust und Ausfällen durch Verweigerung unterschieden¹³.

Operational basiert die Unterscheidung auf dem Vorgehen bei einem Interviewversuch: Jeder Haushalt muss in der folgenden Welle erneut kontaktiert werden. Gelingt dies nicht, wird von *keinem Kontakt* ausgegangen. Allerdings muss dabei eruiert werden, ob Mitglieder des Haushaltes umgezogen sind und falls ja, ob der Umzug ins Ausland erfolgte oder nicht. Schließlich muss geklärt werden, ob eventuell eines oder mehrere der Mitglieder des Haushaltes verstorben sind. Wurde ein Kontakt hergestellt und liegt kein Interview vor, wird von einer *Verweigerung* ausgegangen. Dies hat die Konsequenz, dass auch Perso-

¹¹ Eine Recherche der Datenbanken „Sociological Abstracts“, SOLIS und SRM erbrachte lediglich die Arbeiten der DIW-SOEP-Gruppe.

¹² Zwar finden sich in der Arbeit von Pischner 1994 Gesundheitsvariablen in den Analysetabellen, diese werden im Text der Arbeit aber nicht besprochen.

¹³ Anscheinend wurden die Analysen der SOEP-Arbeitsgruppe bislang nicht unabhängig repliziert. Dies wird durch eine Reihe von Problemen bei den technischen Details der Modelle auch erschwert. So findet sich in der den Autoren bekannten Dokumentation nirgends eine Tabelle, die eine eindeutige Zuordnung der ein- oder zweistelligen Bearbeitungs-codes (\$PERGZ bzw. \$HERGSZ) zu den Ereignissen „Kontaktverlust“ bzw. „Verweigerung“ enthält.

nen, die aus medizinischen Gründen dauerhaft befragungsunfähig werden, in der Analyse als Verweigerungen geführt werden.

Den Arbeiten der SOEP-Arbeitsgruppe zu Panelausfällen kommt aus einem Grund besondere Bedeutung zu: Aus den Logitmodellen zur Vorhersage des „Kontaktverlustes“ einerseits und der „Verweigerung“ andererseits werden die Längsschnittgewichte des SOEP berechnet. Bei allen Analysen dieser Gruppe werden die Ausfälle stets auf der Ebene der Haushalte analysiert, nicht hingegen auf der Personenebene. Auf Einzelheiten diese Modelle wird weiter unten eingegangen.

Vorbereitende Untersuchungen

Im Rahmen der vorbereitenden Untersuchungen zur gesundheitsbedingten Panelmortalität wurden für alle Wellenübergänge die Angaben zu den Gesundheitsvariablen der vorhergehenden Welle in Abhängigkeit von den Bearbeitungsergebnissen in der Folgewelle (\$perg) analysiert¹⁴. Die Ergebnisse ähneln sich über die Erhebungswellen. Als Beispiel werden hier die Ergebnisse für den

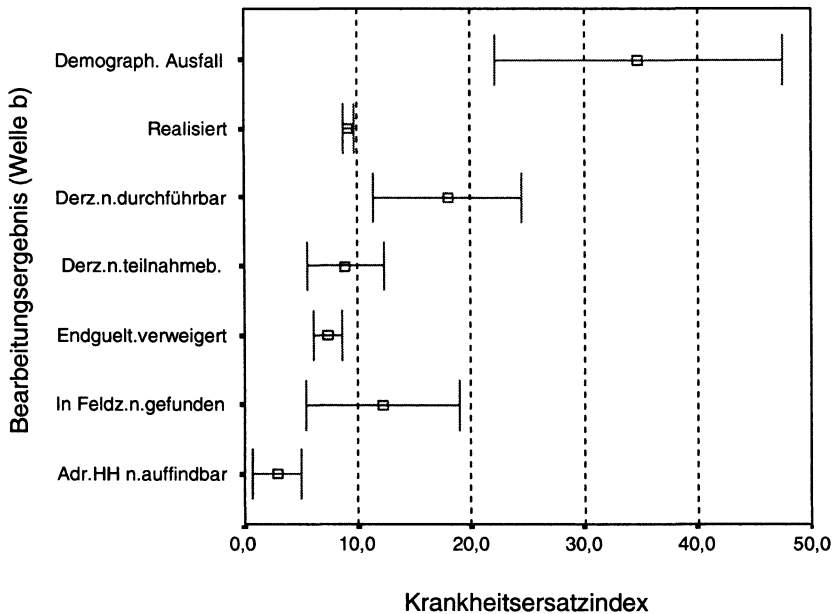


Abb.1: 95%-Konfidenzintervalle für den Krankheitsersatzindex in Welle a in Abhängigkeit vom Bearbeitungsergebnis in Welle b (Westdeutsche)

¹⁴ Die hier berichteten Responsekategorien (=Bearbeitungsergebnisse) entsprechen den im SOEP verwendeten Definitionen der Variablen *bperg*. Alle im Folgenden berichteten Analysen wurden mit ungewichteten Datensätzen durchgeführt.

Krankheitsersatzindex beim Übergang von Welle A zu Welle B dargestellt. Die Abbildung stellt Mittelwerte und die zugehörigen 95% Konfidenzintervalle dar. Sowohl in der graphischen Darstellung, wie auch in einer ANOVA zeigen sich signifikante Mittelwertdifferenzen zwischen den Responsekategorien in Hinsicht auf den KEI ($F=24.02$, $df=6/9060$, $p < 0.0001$, vgl. Abbildung 1). Ähnliche Ergebnisse zeigen sich auch für die Gesundheitszufriedenheit ($F=13.74$, $df=6/9053$, $p < 0.0001$). Interessanterweise sind die Unterschiede bei den Gesundheitsvariablen zwischen den Responsekategorien bei den Ausländern weniger stark ausgeprägt (vgl. Abbildung 2). Dies lässt sich einfach erklären. Betrachtet man z.B. die Zufriedenheit mit der Gesundheit 1984, so zeigt sich in Hinsicht auf die Ausfälle durch Tod bzw. Emigration 1985 ein deutlicher Unterschied zwischen Deutschen und Ausländern. Beschränkt man die Analyse auf die über 60jährigen, so zeigt sich kein signifikanter Unterschied zwischen Ausländern und Deutschen. Die Ausländer in der BRD stellen demnach ein Beispiel für „healthy migrants“ dar: Migranten sind eher jünger und gesünder als Nichtmigranten. Wir beschränken uns in den folgenden Analysen daher auf Deutsche.

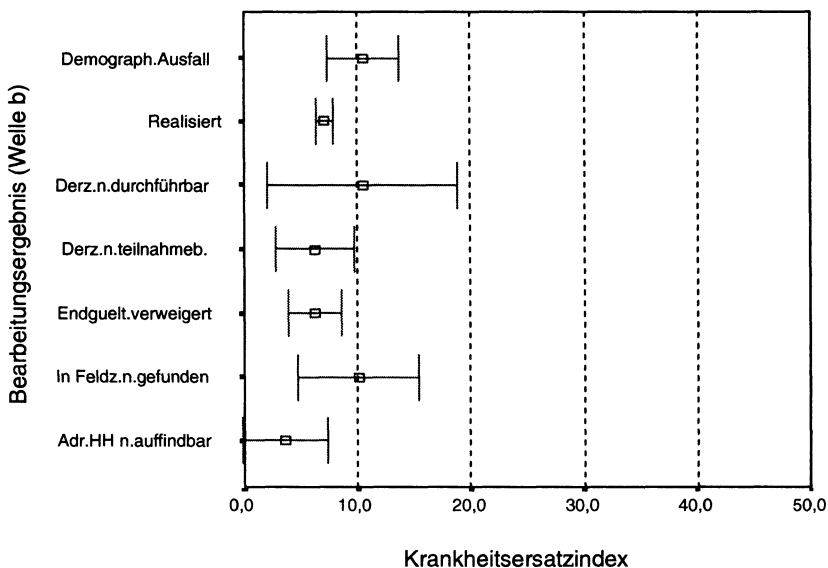


Abb.2: 95%-Konfidenzintervalle für den Krankheitsersatzindex in Welle a in Abhängigkeit vom Bearbeitungsergebnis in Welle b (Ausländer)

Auffällig bei allen Analysen sind insbesondere die unter „demographischer Ausfall“ aufgeführten Teilnehmer (Personen, die ins Ausland gezogen oder verstorben sind)¹⁵. Diese weisen in der vorherigen Welle signifikant höhere

¹⁵ Operational sind das Personen, die in der Datei PFAD in der Variablen *\$netto* die Werte 0 oder -2 besitzen.

KEI-Werte bzw. eine niedrigere Gesundheitszufriedenheit auf. Betrachtet man die Kategorie etwas genauer, so legt der bei den Ausländern beobachtete „healthy migrant“-Effekt die Vermutung nahe, dass auch bei den Deutschen zwischen den beiden Gruppen (Migranten und Verstorbene) unterschieden werden muss. Wiederholt man die Analyse der Abbildung 1 ohne die (mit Sicherheit) 1984 oder 1985 Verstorbenen, so zeigt sich kein solcher Unterschied mehr zwischen den anderen Bearbeitungskategorien. Vielmehr erscheinen die Emigrierten tendenziell gesünder als die Befragten (vgl. Abbildung 3).

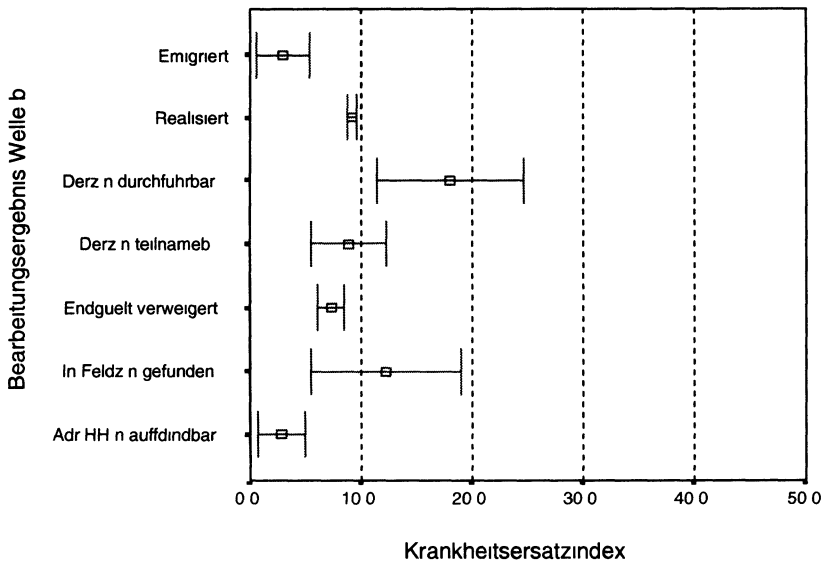


Abb 3 95%-Konfidenzintervalle für den Krankheitsersatzindex in Welle a in Abhängigkeit vom Bearbeitungsergebnis in Welle b (Westdeutsche, ohne Verstorbene 1984-85)

Die Gruppe der demographischen Ausfälle wird zumeist als „unechte“ Panelmortalität aufgefaßt. Daher erfolgen selten Analysen dieser Responsegruppe¹⁶

Die hier berichteten Analysen zusammenfassend kann festgehalten werden:

- den verwendeten Gesundheitsindikatoren „Gesundheitszufriedenheit“ und dem „Krankheitsersatzindex“ kann prädiktive Validität zugesprochen werden.

¹⁶ Bei Analysen des SOEP wird diese Gruppe leicht übersehen. Im Datensatz für das personenbezogene Bearbeitungsergebnis (*\$perg* in *\$PBRUTTO*) liegen keine Daten für diese Personengruppe vor. Die Information für die Art des Ausfalls stammt aus der Variablen *\$netto* in *PPFAD* und muss entsprechend zugespielt werden. Rechnet man mit den personenbezogenen Bearbeitungsergebnissen (was bei einer Non-responseanalyse nahelegt), so fällt diese Gruppe aus den Analysen aufgrund fehlender Werte aus.

den: Die im Übergang zur nächsten Welle Verstorbenen weisen signifikant ungünstigere Werte auf diesen Gesundheitsindikatoren auf.

- es gibt deutliche Unterschiede zwischen der Stichprobe der Westdeutschen und den Ausländern auf den Gesundheitsindikatoren; ein „healthy migrant effect“ kann bestätigt werden
- eine Unterscheidung zwischen Ausfall durch Migration und Ausfall durch Tod ist dringend erforderlich, da der Gesundheitszustand der Migranten zumindest nicht schlechter als der der Befragten ist
- in mehreren Wellen existieren zwischen verschiedenen Responsegruppen (vor allem zwischen „realisiert“ und „Interview derzeit nicht durchführbar“) signifikante Unterschiede in den Gesundheitsvariablen¹⁷.

Vor allem die beiden letzten Punkte zeigen, dass Analysen gesundheitsbedingter Ausfälle nicht mittels Kategorien durchgeführt werden sollten, in denen Ausfälle durch verschiedenste Ursachen zusammengefasst werden. Viel sinnvoller wäre es, sorgfältig recodierte Aufgliederungen der detailliert im SOEP dokumentierten Bearbeitungsergebnisse (die sogenannten „zweistelligen Bearbeitungs-codes“) zu verwenden. Im Folgenden werden wir uns aber auf eine Analyse der Panelmortalität unter besonderer Beachtung bisher geleisteter Vorarbeiten beschränken, die eine solche Unterteilung nicht vorgenommen haben.

Einfluss von Gesundheitsvariablen auf das Gewichtungsmodell des SOEP

Die SOEP-Arbeitsgruppe des DIW verwendet auf Haushaltsebene je ein Logitmodell zur Erklärung des Kontaktverlustes bzw. der Verweigerung. Die durch die Modelle „vorhergesagten“ Wahrscheinlichkeiten werden zur Längsschnittgewichtung des SOEP verwendet¹⁸. Die genannten Arbeiten sind im Detail zum Teil schwer nachvollziehbar. Um einen kritischen Umgang mit solchen Analysen zu ermöglichen, haben wir unser Vorgehen so genau wie möglich dokumentiert.

Da in unseren Analysen der individuelle Gesundheitszustand von Interesse ist, ist hier vor allem die Fähigkeit der Modelle das individuelle Responseverhalten prognostizieren zu können von Bedeutung.

¹⁷ Im Datensatz des SOEP werden in der Variablen *bpergz* so unterschiedliche Angaben wie: „alt und krank“, „Krankenhaus über Feldzeit“, „Ausländer längere Zeit im Heimatland“, „in der Feldzeit nicht erreicht“, „nicht auswertbar“, „sonstiger unklarer Fall“ zu einer einzigen Ausprägung in der Variablen *bperg* (Interview derzeit nicht durchführbar) zusammengefasst.

¹⁸ Diese Art der Gewichtung wird in der neueren Statistikliteratur als „propensity weighting“ bezeichnet, vgl. hierzu Schnell 1997.

Ein Verfahren zur Beurteilung solcher Modelle sind Untersuchungen zum Fit des Gewichtungsmodells. Die Beurteilung des Modellfits von Logitmodellen ist alles andere als trivial (Collett 1991). Neben den üblichen (und umstrittenen) Fitmaßen wie verschiedenen Pseudo-R²-Statistiken, der Anzahl korrekt klassifizierter Fälle, des Hosmer-Lemeshow-Tests etc. werden eine Reihe von Standardtests statistischer Modelle verwendet, die letztlich überprüfen, ob weitere Variablen systematisch mit den Residuen des Modells kovariieren (Hosmer & Lemeshow 1989). Fittet das Modell die Daten, dann sollten weitere Variablen nicht mit den Residuen kovariieren und entsprechend (in der Regel) die Koeffizienten der bereits im Modell enthaltenen Variablen nicht verändern falls diese mit den neuen Variablen kaum zusammen hängen.

Interessanterweise wird von der SOEP-Gruppe für jede Welle ein anderes Modell für die Vorhersage des Kontaktverlustes und für die Vorhersage der Verweigerung angegeben. Diese Modelle unterscheiden sich jeweils durch Hinzufügen oder Weglassen einzelner Variablen bzw. einzelner Interaktionsterme. Es handelt sich also nicht um theoretische Modelle, sondern um empirisch ausgewählte Modelle. Die Kriterien, anhand derer die Modelle ausgewählt wurden, werden nicht eindeutig genannt („For the computation of the GSOEP weighting schemes only model specifications with all covariates being significant were used“, Pannenberg 1998). Die Berechnung und der Nachvollzug der Modelle wird dadurch erschwert, dass in die Definition einzelner Variablen andere Variablen eingehen, die zusätzlich als Bestandteil weiterer Interaktionseffekte Verwendung finden¹⁹.

Für die Überprüfung dieser Modelle sind bei gesundheitsbezogenen Analysen natürlich zunächst Gesundheitsvariablen von Interesse. Im ersten Schritt der Analyse wurde ein an die SOEP-Modell eng angelehntes Modell zur *Vorhersage des Kontaktverlustes* geschätzt. Das Modell enthielt die zuletzt bei Pannenberg (1998) verwendeten Variablen (Umzug des Haushaltes kombiniert mit Haushaltsgröße, Haushaltsgröße, Einpersonenhaushalt, Haustyp, Großstadt und abgespaltener Haushalt) sowie zusätzlich die Zufriedenheit mit der Gesundheit, den Krankheitsersatzindex, die Pflegebedürftigkeit und die Schwerbehinderung. Von der Annahme ausgehend, dass Gesundheit und Krankheit individuelle Merkmale darstellen und das persönliche individuelle Responseverhalten

¹⁹ Ein Beispiel für die Kovariaten eines Logit-Modells zur Schätzung eines „Nicht Kontaktes“ in der Folgewelle stellt die Variable *MOVE* dar, die bei Pannenberg (1998) wie folgt definiert ist:

	<i>MOVE</i> 1	Household not moved
2		Moved multi person household
3		Moved single person household
4		Split off household

Dabei ist gut zu erkennen, dass diese Variable Informationen aus mehreren Variablen (Adressenänderung, Alter, neuer Haushalt und Haushaltgröße) beinhaltet und Interaktionen enthält. Dabei ist es für die Übersichtlichkeit nicht hilfreich, dass später aus dieser Variable wiederum eine Interaktionsvariable (*SINGLE*MOVE*; Pannenberg 1998) gebildet wird.

beeinflussen, wurde hier auf Personenebene (nicht auf Haushaltsebene) analysiert²⁰.

Für alle untersuchten Wellen (A-F) liefern die Pseudo-R²-Werte der Modelle zur Prognose des individuellen Kontaktverlustes meistens befriedigende Werte (0.15 bis 0.47, allerdings für die Welle B nur 0.06)²¹. Da nur wenige der möglichen Kovariaten Gruppen tatsächlich existieren, sind diese Werte eher irreführend. Entsprechend lassen sich Standardtests des Fits (Hosmer-Lemeshow, Pearson) nur für wenige Wellen durchführen. Betrachtet man die Klassifikationstabellen, so zeigt sich über alle Wellen hinweg, dass für alle Fälle immer nur eine Ausprägung („erreicht“) vorhergesagt wird. Der Fit der Modelle zur Vorhersage des Kontakts auf Personenebene ist eher schlecht.

Allerdings zeigt kaum eine der Gesundheitsvariablen einen nachweisbaren Effekt auf die vorhergesagte Wahrscheinlichkeit eines Kontaktverlustes (aus Platzgründen wird auf eine detaillierte Darstellung dieser Modelle an dieser Stelle verzichtet)²². Eine wesentliche Ausnahme stellt allerdings die Variable Pflegebedürftigkeit dar. Bis auf eine Person in Welle d werden alle Befragten, für die in der vorherigen Welle eine Pflegebedürftigkeit festgestellt wurde in der folgenden Welle wieder erfolgreich kontaktiert.

Zusammenfassend lässt sich für die Vorhersage des Kontaktverlustes festhalten, dass sich dieser durch die Modelle kaum erfolgreich vorhersagen lässt. Allerdings scheinen erfreulicherweise auch die meisten Gesundheitsindikatoren keinen Effekt auf die Wahrscheinlichkeit eines Kontaktverlustes zu besitzen.

Für die *Vorhersage der Verweigerung* verwendet das SOEP-Modell weit mehr Variablen, z.B. Interviewerwechsel, Alter des Haushaltsvorstandes, Arbeitslosigkeit des Haushaltsvorstandes, neue Adresse des Haushaltes, Haushaltseinkommen, Anzahl der Vermögensarten, keine Angabe des Haushaltseinkommens. Auch hier wird von Welle zu Welle ein anderes Modell angepasst. Die Begründung der Auswahl der jeweiligen Prädiktoren erfolgt durch das gleiche Argument wie beim Kontaktmodell. Die Tabelle 1 illustriert auszugsweise die Variablendefinition, die Tabelle 2 die geschätzten Logitmodelle für Verweigerungen in der Westdeutschen Stichprobe.

²⁰ Rendtel (1993) zufolge determiniert das Responseverhalten des Haushaltes (Haushaltsvorstandes) im wesentlichen das Teilnahmeverhalten der Haushaltsmitglieder. Folglich können die genannten Modelle der SOEP-Arbeitsgruppe auch zur Analyse von individuellem Responseverhalten genutzt werden.

²¹ Die Merkmale auf Haushaltsebene, z.B. „Alter des Haushaltsvorstandes“ oder „Haushaltsnettoeinkommen“, wurden den Personen (deren Zugehörigkeit zu einem Haushalt entsprechend) zugewiesen.

²² Folgt man aber der Logik des Ursprungsmodells und lässt in verschiedenen Wellen unterschiedliche Dummy-Variablen, bzw. Rekodierungen der Gesundheitsindikatoren zu, lassen sich durchaus signifikante Effekte für die Gesundheitsvariablen erzielen.

Characteristic	Abbreviation	Code	Values
Age of the head of household	ALTHV	1 2 3 4 5 6	Older than 75 years 65-74 years 55-64 years 35-54 years 25-34 years Younger than 25 years
Gender of the head	SEX	0 1	Male Female
Typ of the household	HTYP	1 2 3	Old household without move Old household moved Split-off household
Change of Interviewer	INTW	0 1 2	No change Change since previous wave Not regular interviewer number
Number of interviews	BETREUUNG		Number of interviews with the interviewer of the present wave
Household income West-Germany	EINKW	1 2 3 4	Income not reported < 2000 DM 2000-4000 DM 4000 DM
Household income not reported	KAEINK	0 1	Else Income not reported
Balance of assets not reported	KAVB	0 1	Else Balance not reported in wave 5
Number of different kinds of assets in the households	ANZAHL	1 2 3	Number=0 Number=5 (Maximum) Else
No assets reported	ANZ0	0 1	Else Number of reported assets=0
Firm assets	BETRIEB	0 1	Else Household owns firm assets

Tabelle 1: Variablenliste Vorhersage der Verweigerungen, Wellenübergänge 1-5, Westdeutschenstichprobe (gekürzt nach Pannenberg 1998)

Bei der Prüfung des Einflusses der Gesundheitsvariablen auf das „Verweigerungsmodell des SOEP“ wurde analog wie bei der *Vorhersage des Kontaktverlustes* vorgegangen. Es wurden die Variablen des SOEP-Verweigerungsmodells benutzt. Allerdings wurde das Responseverhalten personenbezogen analysiert. Eigenschaften der Haushalte wurden den Haushaltsmitgliedern (Personen) zugewiesen. Auf Grund modellimmanenter Schätzprobleme wurde, abweichend von den DIW Modellen, auf eine (exzessive) Nutzung von Interaktionsvariablen verzichtet (vgl. Fußnote 24).

Subsample A (West-Germans)

Wave	Model and coefficients
2	Model=CONST+INTW+ALTHV+HTYP+EINKW CONST (-1.53), INTW (0: -0.25 / 1: 0.25), ALTHV (1: 0.66 / 2,3,4: 0.03 / 5: -0.39 / 6: -0.30), HTYP (1: -0.68 / 2: -0.19 / 3: 0.87), EINKW (1: 0.61 / 2: 0.12 / 3: -0.35 / 4: -0.38)
3	Model=CONST+INTW+ALTHV+INTW*ALTHV+HTYP+ALOS+KAEINK CONST (-1.22), INTW (0: -0.39 / 1: 0.39), ALTHV * (INTW=0) (1: -0.13 / 2: -0.11 / 3,4: -0.39 / 5: 0.26 / 6: 0.37), ALTHV * (INTW=1) (1: 0.13 / 2: 0.11 / 3,4: 0.39 / 5: -0.26 / 6: -0.37), ALTHV (1: 0.59 / 2: 0.16 / 3,4: -0.06 / 5: -0.53 / 6: -0.16) HTYP (1: -0.52 / 2: 0.10 / 3: 0.42), ALOS (0: -0.21 / 1: 0.21), KAEINK (0: -0.39 / 1: 0.39)
4	Model= CONST+ALTHV+INTW (ALTHV) + HTYP + KAEINK CONST (-1.83), INTW (ALTHV=1) (0: -0.44 / 1: 0.44), INTW (ALTHV=2) (0: -0.74 / 1: 0.74), INTW (ALTHV= 3,4) (0: -0.59 / 1: 0.59), INTW (ALTHV= 5) (0: -0.41 / 1: 0.41), INTW (ALTHV= 6) (0: -0.32 / 1: 0.32) ALTHV (1: 0.21 / 2: -0.38 / 3,4: -0.24 / 5: 0.06 / 6: 0.35), HTYP (1: -0.45 / 2: 0.29 / 3: 0.19), KAEINK (0: -0.39 / 1: 0.39)
5	Model=CONST+BETREUUNG+ALTHV (INTW=1) + HTYP +KAEINK + ANZ0 Const (-1,60) BETREUUNG (1: 1.15 / 2: 0.41 / 3: 0.18 / 4: -0.71 / 5: -1.03), ALTHV (INTW=1) (1,2: 0.52 / 3,4,5: -0.11 / 6: -0.40), HTYP (1: -0.49 / 2: 0.11 / 3: 0.38), KAEINK (0: -0.45 / 1: 0.45), ANZ0 (0: -0.38 / 1: 0.38)
6	Model=CONST+BETREUUNG+ALTHV (INTW=1)+ HTYP + KAEINK + KAVB + BETRIEB CONST (-2.44), BETREUUNG (1: 0.75 / 2: 0.58 / 3: 0.21 / 4: -0.59 / 5: -0.43 / 6: - 0.52), ALTHV (INTW=1) (1,2: 0.26 / 3,4,5 0.05 / 6: -0.31), HTYP (1: -0.32 / 2: -0.04 / 3: 0.37), KAEINK (0: -0.26 / 1: 0.26), BETRIEB (0: 0.41 / 1: -0.41)

Tabelle 2: Logitmodelle: Vorhersage der Verweigerungen, Wellenübergänge 1-5, Westdeutschenstichprobe (gekürzt nach Pannenberg 1998)

Betrachtet man die Modelle zur Vorhersage der Verweigerung auf Personenebene, so liegen die Pseudo-R²-Werte eher niedrig (zwischen 0.05 und 0.09). Allerdings ergeben die Hosmer-Lemeshow-Tests meist nicht signifikante Abweichungen zwischen erwarteten und beobachteten Häufigkeiten, so dass der Test den Modellfit eher günstig erscheinen lässt. Führt man eine Regressionsdiagnostik für diese Modelle durch, so ist die relativ geringe Zahl der unterschiedlichen Kombinationen der unabhängigen Variablen auffällig. Entsprechend finden sich bedeutsame Abweichungen der Residuen von der Normalverteilung. Weiterhin besitzen einige wenige Kombinationen der unabhängigen Variablen sehr großen Einfluss auf die Fit-Statistiken und die Parameterschätzungen.

Fügt man in diese Modelle die Gesundheitsindikatoren ein, so verbessern sich die Pseudo-R²-Werte nicht (vgl. Tabelle 3). Allerdings lassen sich für einzelne Gesundheitsindikatoren in einzelnen Wellen durchaus signifikante Effekte demonstrieren. So ist der Effekt der Behinderung bei Übergang von Welle A auf B und von Welle B auf C positiv, d.h. die Behinderten verweigern eher. Für den Wellenübergang D-E zeigt sich dagegen ein umgekehrter Zusammenhang.

Als letzter Schritt der Analyse wurde hier das Versterben einer Person 1-3 Jahre nach dem Wellenübergang als zusätzliche Variable in das logistische Regressionsmodell mit aufgenommen (vgl. Tabelle 4)²³. Betrachtet man z.B. den Übergang von Welle B zu Welle C, dann wird das später bekanntgewordene Versterben der Person nach Welle C als Prädiktor für den Ausfall in Welle C verwendet. Die Logik dieses Vorgehens beruht auf der Annahme, dass der Gesundheitszustand von Panelteilnehmern 1-3 Jahre vor ihrem Tod eher schlechter ist. Aufgrund dieses Gesundheitszustands könnten entweder sie selbst oder andere Haushaltsmitglieder stellvertretend für sie die Teilnahme verweigern²⁴.

²³ Im Rahmen der von Infratest 1992 durchgeführten Verbleibestudie wurde unter anderem nachgeforscht, ob in der Stichprobe befindliche Personen verstorben waren.

²⁴ Diese Analyse steht vor einem praktischen Problem: Die Information über den Tod eines Panelteilnehmers steht nicht für alle Personen zur Verfügung. In der Regel fehlt diese Information für ca. 20% der Westdeutschenstichprobe, da diese Personen irgendwann vor der jeweiligen Welle aus dem Panel ausschieden. Für diese ist nicht bekannt, ob sie verstarben oder nicht. Man kann nun entweder diese Personen aus den Analysen ausschließen, sie zu den Verstorbenen bzw. zu den Lebenden zählen oder als weitere Kategorie aufnehmen. Im letzten Fall ist eine deutliche Modellverbesserung in allen analysierten Wellen feststellbar, die aber auf einem Artefakt beruht: Diejenigen Teilnehmer, über deren Vitalstatus nichts bekannt ist, sind per definitionem ausgefallen. Um solche Artefakte zu vermeiden, wurden diese Personen daher hier aus den Analysen ausgeschlossen. Hierdurch verändert sich aber die Menge der Personen, für die Analysen gerechnet werden. Schließt man nur die entsprechenden Personen aus den ursprünglichen Modellen aus ohne neue Variablen mit ins Modell zu nehmen, verändern sich die Parameterschätzungen bei einigen Variablen (z.B. Betriebsvermögen, keine Angabe Vermögensbilanz etc.) wesentlich. Dies scheint auf eine Fehlspezifizierung des Ausgangsmodells hinzudeuten: Würde das Modell die Ausfälle erklären können, dürfte eine solche Veränderung der Stichprobe nicht solche Veränderungen der Parameterschätzungen bedingen.

Bis auf den Wellenübergang C-D belegen die vorgelegten Analysen für alle untersuchten Wellen einen signifikanten Einfluss des Prädiktors „Tod nach der Erhebungswelle“ auch bei Kontrolle anderer Variablen auf die Wahrscheinlichkeit für eine Verweigerung des Interviews²⁵. Der Effekt dieser Variablen ist in der Regel einer der stärksten im Modell, weiterhin verändern sich die Effektstärken der anderen Variablen nach Einführung dieses Prädiktors häufig. Dass die Variable „Tod nach der Erhebungswelle“ wesentlich stärkere Zusammenhänge mit Verweigerung aufweist als die anderen im Modell befindlichen Gesundheitsvariablen mag als Hinweis darauf verstanden werden, dass diese Variable eher in der Lage ist objektive Gesundheitsveränderungen zu beschreiben als es die übrigen analysierten Gesundheitsvariablen vermögen. Sollte die hier formulierte Hypothese über den Wirkungsmechanismus dieses Prädiktors korrekt sein, dann stellen die beobachteten Daten des SOEP für jede gegebene Welle eine systematisch selektierte Stichprobe dar. Damit würde der Gesundheitszustand der Grundgesamtheit überschätzt. Die bisherige Gewichtung des SOEP kann diese Effekte nicht auffangen. Weitere detailliertere Studien der berichteten Effekte erscheinen für die Analyse gesundheitsbezogener Ausfälle notwendig.

Es muss darauf hingewiesen werden, dass die Daten der Version GSOEP14 von der SOEP-Arbeitsgruppe so umkodiert wurden, dass spätestens ab dieser Version die hier durchgeführten Analysen so nicht mehr möglich sind. Im wesentlichen wurde die Information der „fehlenden Information hinsichtlich des Todes ausgeschiedener Panelteilnehmer“ auf „trifft nicht zu“ umkodiert. Eine Rückfrage beim DIW (Berlin) ergab, dass diese Umkodierung auf keine neuen empirischen oder theoretischen Erkenntnisse zurückzuführen ist, sondern lediglich „aufgegeben“ wurde. Wir haben daher für die hier vorgelegten Analysen auf die informationsreichere Version GSOEP12 zurückgegriffen.

²⁵ Bei vielen Analysen mit diesen Modellen ergibt sich das Problem, dass bestimmte Kombinationen der unabhängigen Variablen die abhängige Variable perfekt vorhersagen. Obwohl dies für die tatsächliche „Prognose“ eines Ausfalls durchaus wünschenswert ist, führt dies häufig zu Problemen bei der Schätzung von Parametern (insbesondere für die hohe Zahl von Interaktionsvariablen) und die damit verbundenen Signifikanztests, die sich bei mehreren Wellen dann nicht berechnen lassen. Aus den hier berichteten Analysen wurden daher die Interaktionsvariablen herausgenommen. Die Effekte für die Variable „Tod“ verändern sich dadurch kaum.

	Welle a-b		Welle b-c		Welle c-d		Welle d-e		Welle e-f	
	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert
Betreuung 1 (1)							1.74	0.19	1.77	0.26
Betreuung 2 (1)							1.35	0.08	0.91	0.64
Betreuung 3 (1)							1.12	0.46	1.25	0.22
Betreuung 4 (1)							1.48	0.00	1.33	0.07
Betreuung 5 (1)									1.25	0.11
Wechsel des Interviewers	1.52	0.00	1.47	0.00	1.39	0.03	0.66	0.29	0.68	0.43
Alter Haushaltsvorstand < 25	0.67	0.06	1.05	0.82	1.66	0.02	1.01	0.97	0.59	0.05
Alter Haushaltsvorstand 25-34	0.81	0.07	0.75	0.03	1.10	0.50	0.98	0.90	0.53	0.00
Alter Haushaltsvorstand 55-64	1.34	0.00	1.24	0.07	0.91	0.52	0.81	0.14	0.89	0.40
Alter Haushaltsvorstand 65-74	1.44	0.00	1.27	0.10	0.74	0.14	0.78	0.17	0.77	0.15
Alter Haushaltsvorstand >75	2.30	0.00	1.74	0.00	0.94	0.80	0.70	0.12	1.07	0.72
Neue Adresse des Haushaltes	1.55	0.00	1.69	0.00	2.19	0.00	1.97	0.00	1.66	0.01
Neuer Haushalt	7.81	0.00	3.20	0.00	3.51	0.00	4.04	0.00	3.76	0.00
Haushaltsvorstand Arbeitslos			1.50	0.00						
Keine Angabe HH-Einkommen	3.71	0.00	2.51	0.00	3.21	0.00	2.60	0.00	2.17	0.00
Haushaltseinkommen ≤ 2000 DM	1.51	0.00								
Haushaltseinkommen ≥ 4000 DM	1.01	0.94								
Anzahl der Vermögensarten							1.96	0.00		
Vermögensbilanz angegeben									1.90	0.01
Betriebsvermögen									0.52	0.01
Gesundheitszufriedenheit	1.01	0.52	0.96	0.02	0.94	0.01	1.01	0.79	1.00	0.87
Krankheitersatzindex	0.997	0.29	0.999	0.75	0.001	0.76	0.999	0.57	0.997	0.35
Pflegebedürftigkeit	-		1.30	0.41	1.36	0.44	1.38	0.44	1.29	0.57
Schwerbehinderung	1.32	0.03	1.24	0.00	(2)		0.87	0.43	1.25	0.22
Pseudo r^2	0.04		0.03		0.04		0.04		0.02	
L2 χ^2 (df)	235.7 (14)		122.24 (14)		119.09(12)		131.52 (18)		71.52 (20)	
N	8796		8019		7793		7593		7283	
(1)	Längsschnittbetrachtung des Interviewerwechsels: Anzahl der Interviews, die der Interviewer der aktuellen Welle mit der Befragungsperson durchgeführt hat (inklusive aller vorhergehenden Wellen)									
(2)	in dieser Welle nur für 478 Personen erhobene Frage, daher nicht ausgewertet									

Tabelle 3: Logistische Regression. Vorhersage der Verweigerungen; Längsschnittgewichtungsvariablen und Gesundheitsvariablen (Westdeutsche)

	Welle a-b		Welle b-c		Welle c-d		Welle d-e		Welle e-f	
	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert	Odds-Ratio	p-Wert
Betreuung 1 (1)							1.73	0.64	1.15	0.91
Betreuung 2 (1)							1.04	0.94	0.58	0.34
Betreuung 3 (1)							0.96	0.93	0.55	0.27
Betreuung 4 (1)							0.57	0.17	1.04	0.92
Betreuung 5 (1)									0.75	0.41
Wechsel des Interviewers	1.43	0.17	1.80	0.12	1.81	0.20	0.70	0.75	0.71	0.77
Alter Haushaltsvorstand < 25	1.09	0.89	1.02	0.97	1.48	0.53	1.63	0.39	1.30	0.62
Alter Haushaltsvorstand 25-34	0.34	0.06	0.31	0.07	1.40	0.46	1.94	0.06	1.05	0.88
Alter Haushaltsvorstand 55-64	1.17	0.66	0.73	0.51	0.87	0.80	1.04	0.90	0.56	0.16
Alter Haushaltsvorstand 65-74	1.31	0.51	1.47	0.40	0.57	0.47	0.70	0.55	0.61	0.32
Alter Haushaltsvorstand >75	1.78	0.18	0.55	0.34	0.96	0.95	0.71	0.59	1.26	0.60
Neue Adresse des Haushaltes	2.99	0.01	2.67	0.06	5.33	0.00	3.37	0.01	2.86	0.01
Neuer Haushalt	10.39	0.00	8.19	0.00	14.30	0.00	6.05	0.02	7.47	0.01
Haushaltsvorstand Arbeitslos			2.12	1.73						
Keine Angabe HH-Einkommen	7.51	0.00	0.69	0.63	0.52	0.53	1.40	0.59	1.19	0.74
Haushaltseinkommen ≤ 2000 DM	1.92	0.04								
Haushaltseinkommen ≥ 4000 DM	1.52	0.25								
Anzahl der Vermögensarten							1.27	0.56		
Keine Vermögensbilanz angegeben									3.24	0.12
Betriebsvermögen									0.51	0.26
Gesundheitszufriedenheit	1.01	0.87	1.06	0.42	0.91	0.22	0.99	0.88	0.95	0.37
Krankheitsersatzindex	1.005	0.23	1.000	0.96	0.999	0.30	0.994	0.43	0.991	0.33
Pflegebedürftigkeit	-		3.51	0.05	3.85	0.09	0.77	0.81	1.17	0.85
Schwerbehinderung	1.29	0.50	1.65	0.33	(2)		1.01	0.98	1.61	0.31
Tod 1-3 Jahre nach aktueller Welle	4.64	0.00	11.89	0.00	3.22	0.07	6.76	0.00	4.62	0.00
Pseudo r ²	0.09		0.11		0.10		0.07		0.06	
L2 chi2 (df)	71.28 (15)		60.03 (15)		48.55 (13)		44.09 (19)		47.72 (21)	
N	5617		5578		5812		5906		5987	
(1) Längsschnittbetrachtung des Interviewerwechsels: Anzahl der Interviews, die der Interviewer der aktuellen Welle mit der Befragungsperson durchgeführt hat (inklusive aller vorhergehenden Wellen)										
(2) in dieser Welle nur für 478 Personen erhobene Frage, daher nicht ausgewertet										

Tabelle 4: Log. Reg. Vorhersage der Verweigerungen: Längsschnittgewichtungsvariablen, Gesundheitsvar, und Tod 1-3 Jahre nach Wellenübergang (Westdeutsche)

Zusammenfassung

Die hier berichteten Analysen lassen sich folgendermaßen zusammenfassen:

- Sowohl die Zufriedenheit mit dem Gesundheitszustand als auch der Krankheitsersatzindex besitzen Kriteriumsvalidität in Hinsicht auf die Prognose des Ausfalls in späteren Wellen durch Tod des Befragten.
- Sowohl für Ausländer als auch für Deutsche lässt sich ein „healthy migrant“-Phänomen belegen.
- Insbesondere für gesundheitsbezogenen Analysen sollten die Responsekategorien anders zusammengefasst werden als bislang üblich. Die übliche Klassifikation mischt eine Reihe sehr heterogener Prozesse, die besser getrennt analysiert werden sollten.
- Weder das Kontaktverlustmodell noch das Verweigerungsmodell erlauben zufriedenstellende individuelle Prognosen des Ausfalls.
- In den hier analysierten Wellenübergängen führt Pflegebedürftigkeit in der vorherigen Welle (abgesehen von einer einzigen Person) zu einem erfolgreichen Kontakt in der Folgewelle. Für die übrigen gesundheitsbezogenen Variablen ist kein Einfluss auf die Wahrscheinlichkeit eines Kontaktverlustes feststellbar.
- Zwischen der Wahrscheinlichkeit einer Verweigerung und einigen Gesundheitsvariablen lässt sich bei einigen Wellen ein signifikanter Zusammenhang zeigen.
- Ein Indikator der objektiven gesundheitlichen Beeinträchtigung (Tod 1-3 Jahre nach dem Interview) hat einen starken Effekt auf die Verweigerung einer vorherigen Welle. Vermutlich schwer erkrankte Personen verweigern eher bzw. für diese wird stellvertretend verweigert.

Dass dieser Indikator unter Kontrolle der übrigen Gesundheitsvariablen einen so starken Effekt zeigt mag zunächst als ein indirekter Hinweis darauf verstanden werden, dass er objektive Gesundheit/Krankheit besser abzubilden vermag. Da aber davon ausgegangen werden kann, dass offensichtlich moribunde Teilnehmer objektiv kränker sind als beispielsweise solche, die nur überdurchschnittliche Unzufriedenheit ihrer Gesundheit angegeben haben, entpuppt sich dieser starke Zusammenhang als ein überaus deutlicher Hinweis auf eine relevante gesundheitsbedingte Panelmortalität in den ersten 5 Wellenübergängen des SOEP.

Sollte diese Interpretation der berichteten Daten korrekt sein, dann stellen die beobachteten Daten des SOEP für jede gegebene Welle eine systematisch selektierte Stichprobe dar. Damit würde der Gesundheitszustand der Grundgesamtheit überschätzt. Die bisherige Gewichtung des SOEP kann diese Effekte nicht auffangen. Weitere detailliertere Studien der berichteten Effekte an ande-

ren Wellen erscheinen für die Analyse gesundheitsbezogener Ausfälle notwendig.

Literatur

- Cheesbrough S (1993). Characteristics of non-responding households on the Family Expenditure Survey. *Survey Methodology Bulletin* 1993 33: 12-19.
- Cochrane WG (1972). Stichprobenverfahren. Berlin: de Gruyter.
- Collett D (1991). *Modelling Binary Data*. London: Chapman & Hall.
- Engel U, Pötschke M (1997). Erwerbsbiographien in Deutschland: Eine Panelanalyse zur Stabilität von Arbeitslosigkeitskarrieren. In: Klein G, Strasser H (Hrsg.). *Schwer vermittelbar. Zur Theorie und Empirie der Langzeitarbeitslosigkeit*. Opladen: Westdeutscher Verlag.
- Engel U, Reinecke J (1994). *Panelanalyse, Grundlagen Techniken Beispiele*. Berlin: Walter de Gruyter.
- Esser H, Grohmann H, Müller W, Schäffer KA (1989). *Mikrozensus im Wandel, Untersuchungen und Empfehlungen zur inhaltlichen und methodischen Gestaltung*. Band 11 der Schriftenreihe Forum der Bundesstatistik. Stuttgart: Metzler Poeschl.
- Fuchs J, Hansmeier T (1996). Ein Krankheitsersatzindex: Konstruktion und Validierung. *Sozial- und Präventivmedizin*: 41: 231-239.
- Hanefeld U (1987). *Das Sozio-ökonomische Panel. Grundlagen und Konzeption*. Frankfurt/Main: Campus.
- Hosmer DW & Lemeshow S (1989). *Applied Logistic Regression*. New York: Wiley & Sons.
- Kasprzyk D, Duncan G, Kalton G, Sing MP (eds.)(1989). *Panel surveys*. New York: Wiley & Sons.
- Moum T (1994). Is subjective well being a predictor of nonresponse in broad population surveys. *Social Indicators Research* 32: 1-20.
- Pannenberg M (1997). *Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (GSOEP) (1984 until 1996)*. Deutsches Institut für Wirtschaftsforschung, DIW-Diskussionspapier, No 150.
- Pannenberg M (1998). *Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (GSOEP) (1984 until 1997)*. Deutsches Institut für Wirtschaftsforschung, DIW-Diskussionspapier, No 172.
- Pannenberg M, Pischner R, Rendtel U, Wagner GG (1998) *Sampling and Weighting*. Chapter 4 in: Haisken-De New JP, Joachim R. Frick (Eds.) *DTC Desktop Companion to the German Socio-Economic Panel Study (GSOEP) Version 2.2 - September 1998*.
- Pannenberg M, Rendtel U (1996). *Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (GSOEP) (1984 until 1995)*. Deutsches Institut für Wirtschaftsforschung, DIW-Diskussionspapier, No. 137.
- Pischner R (1994). *Quer- und Längsschnittgewichtung des Sozio-ökonomischen Panels (SOEP)*. In: Gabler S, Hoffmeyer-Zlotnik JHP, Krebs D (Hrsg) *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag.

- Pol FJR van de (1989). Reducing panel bias. A review of sampling designs. Chapter 2. In: Pol FJR van de, Langeheine R, Leeuw J de (Eds) Issues of design and analysis of panels. Amsterdam: Sociometric Research Foundation.
- Rendtel U (1995a). Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität. Frankfurt/Main: Campus.
- Rendtel U (1995b). Dokumentation des Sozio-ökonomischen Panels (SOEP). Erhebungsdesign, Fallzahlen und erhebungsbedingte Ausfälle sowie die Schätzung von Ausfallwahrscheinlichkeiten bis Welle 11 (1984 bis 1994). Deutsches Institut für Wirtschaftsforschung Diskussionspapier Nr. 110.
- Schnell R (1991). Wer ist das Volk? Zur faktischen Grundgesamtheit bei allgemeinen Bevölkerungsumfragen. Kölner Zeitschrift für Soziologie und Sozialpsychologie 43: 106-137.
- Schnell R (1993). Die Homogenität sozialer Kategorien als Voraussetzung für „Repräsentativität“ und Gewichtungsverfahren. Zeitschrift für Soziologie 22: 16-32.
- Schnell R (1997). Nonresponse in Bevölkerungsumfragen; Ausmaß, Entwicklung und Ursachen. Opladen: Leske und Budrich.
- Schnell R, Hill P, Esser E (1999). Methoden der empirischen Sozialforschung. München, 6. Auflage: Oldenbourg Verlag.
- Soep-Info (1998)
<http://www-soep.diw-berlin.de/~sieber/soepinfo3.0b/soepinfo3.0b.html>
 Updated to Wave 14 (N). Berlin: German Institute for Economic Research (DIW).